U.S. DEPARTMENT OF ENERGY

INCITE

LEADERSHIP COMPUTING

INCITE Proposal Writing Tips

January 24, 2011

Katherine Riley,
ALCF Acting Deputy Science Director
Argonne National Laboratory
and

Hai Ah Nam,
OLCF Scientific Computing Group
Oak Ridge National Laboratory

Outline

- INCITE program overview
- User's view of systems
- Walk through the proposal form
- Review and awards process
- Links and contacts

What is INCITE?

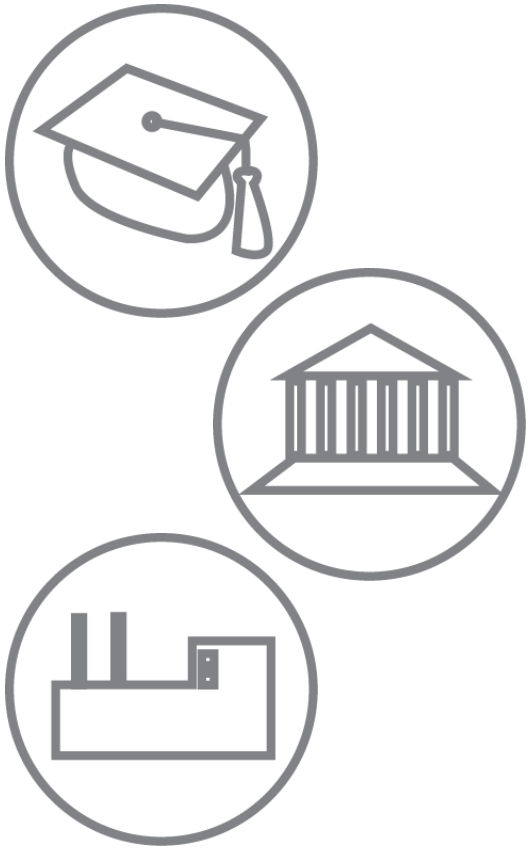
INCITE: Innovative and Novel Computational Impact on Theory and Experiment

Provides awards to academic, government, and industry organizations worldwide needing large allocations of computer time, supporting resources, and data storage to pursue transformational advances in science and industrial competitiveness.

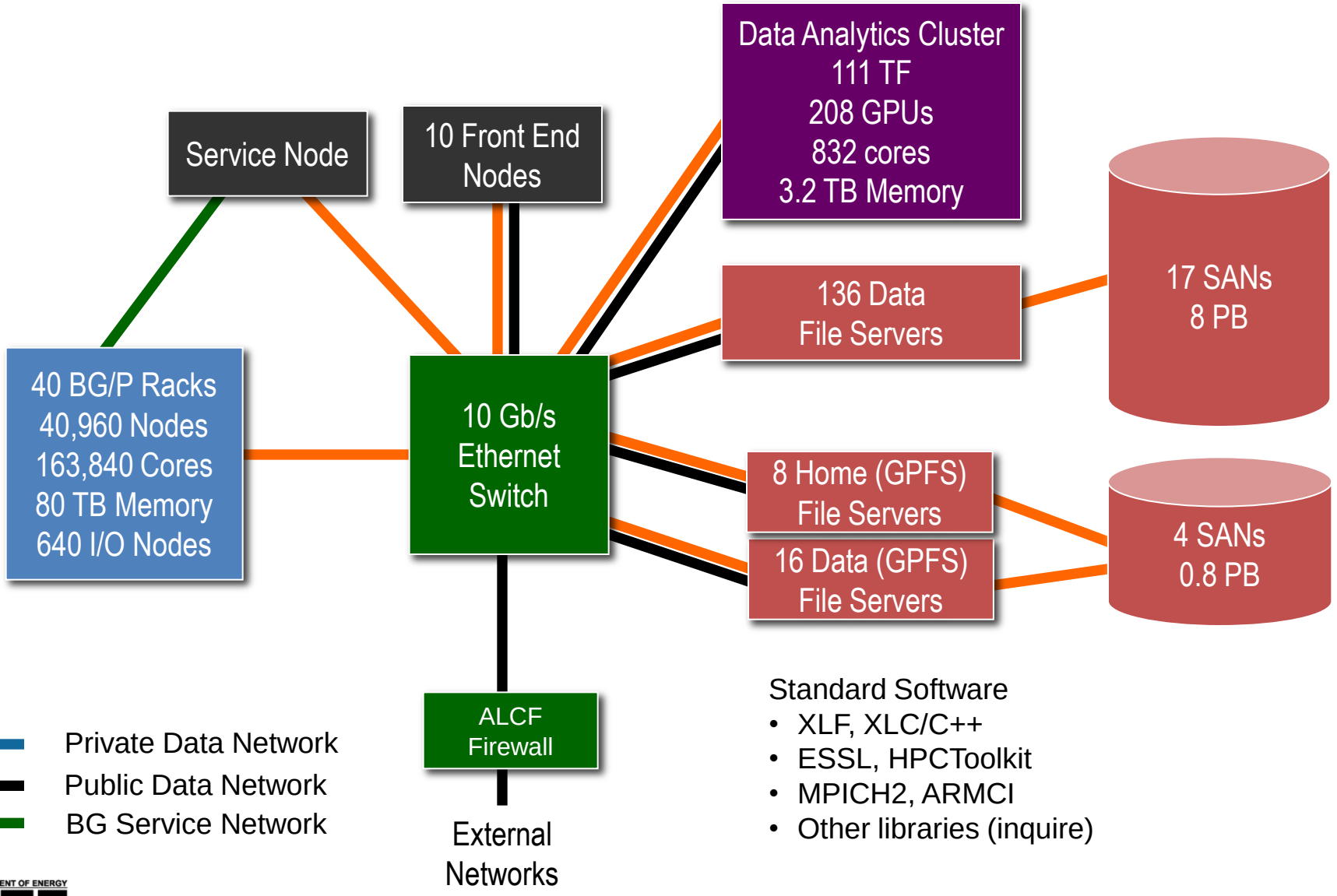
Beginning in 2010, INCITE is jointly run by the ALCF and OLCF, managed by Julia White

INCITE
Innovative and Novel Computational Impact on Theory and Experiment

- Solicits large-scale, computationally intensive research projects
- Open to all scientific & engineering researchers and organizations worldwide
- Provides large computer time and data storage allocations on
  - ALCF IBM BlueGene/P “Intrepid”
  - OLCF Cray XT5 “Jaguar”



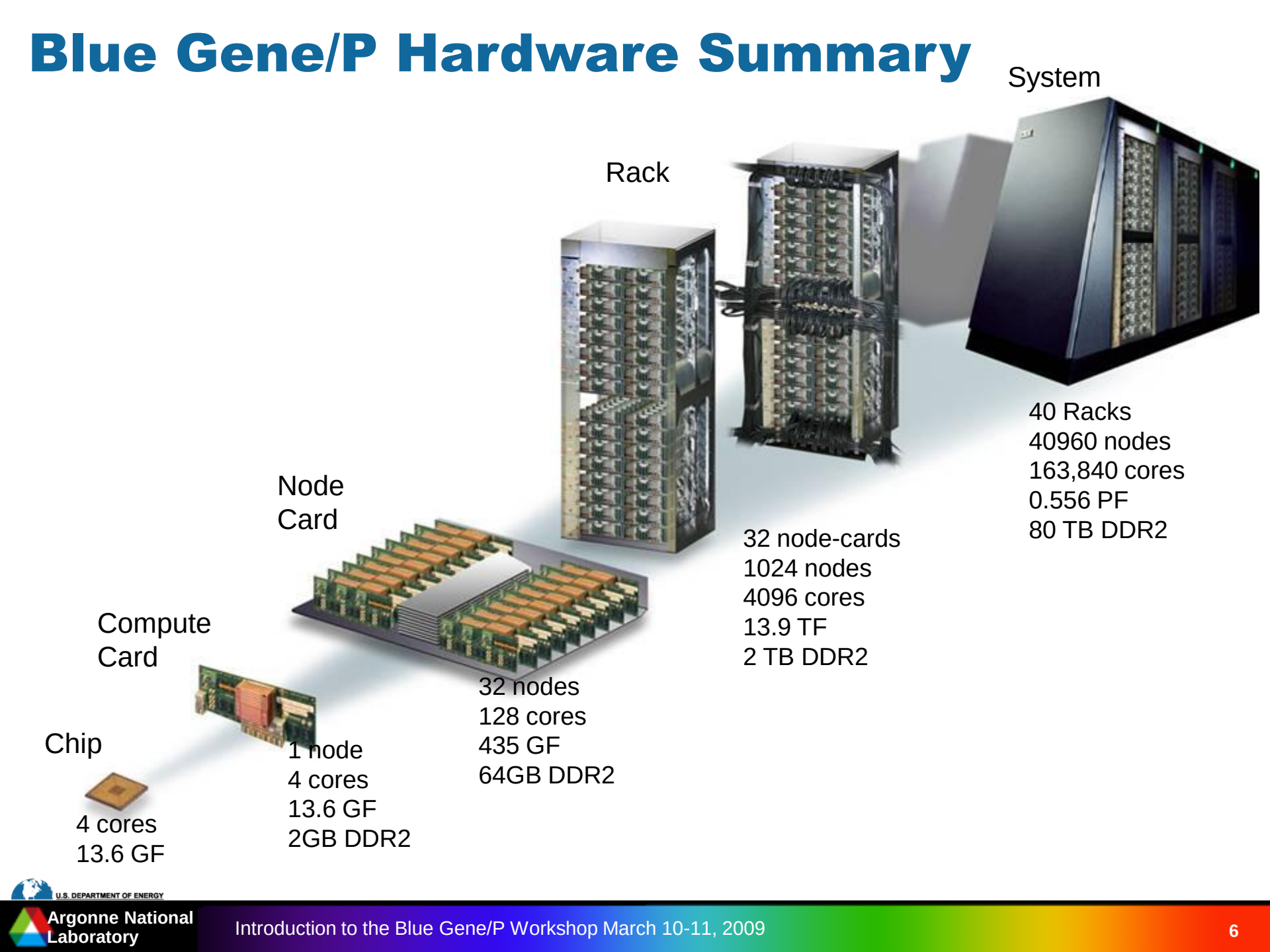
ALCF Blue Gene/P "Intrepid" System



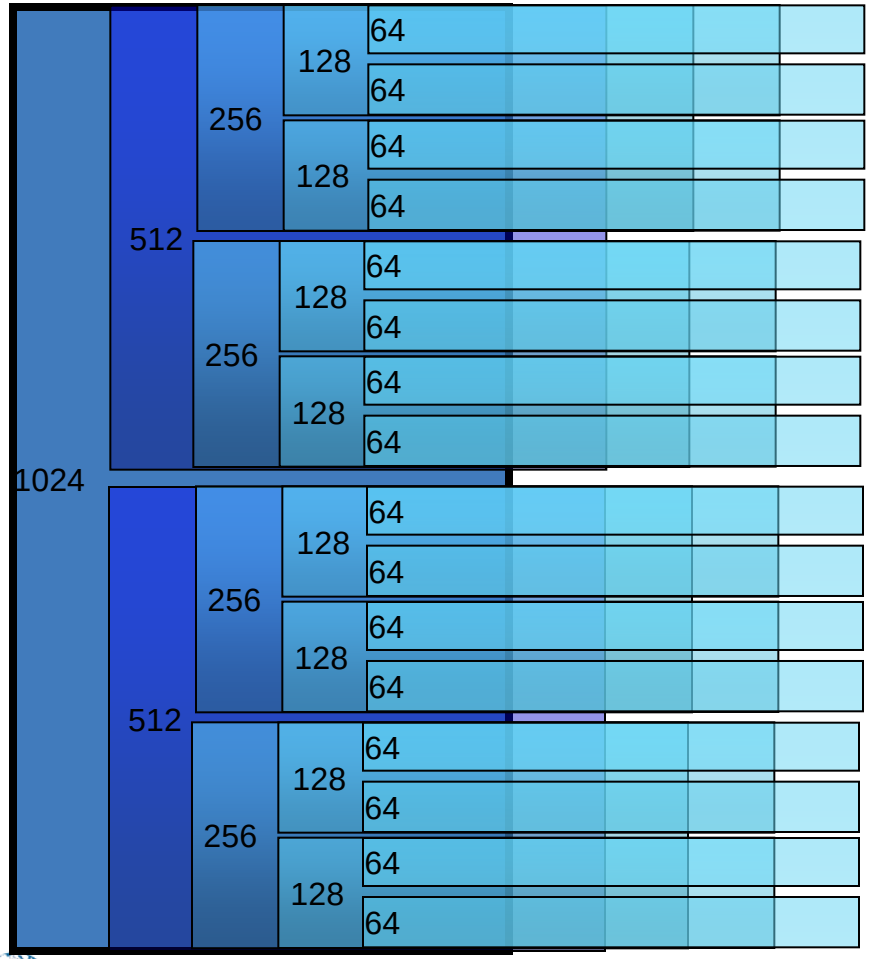
Standard Software

- XLF, XLC/C++
- ESSL, HPCToolkit
- MPICH2, ARMCI
- Other libraries (inquire)

Blue Gene/P Hardware Summary



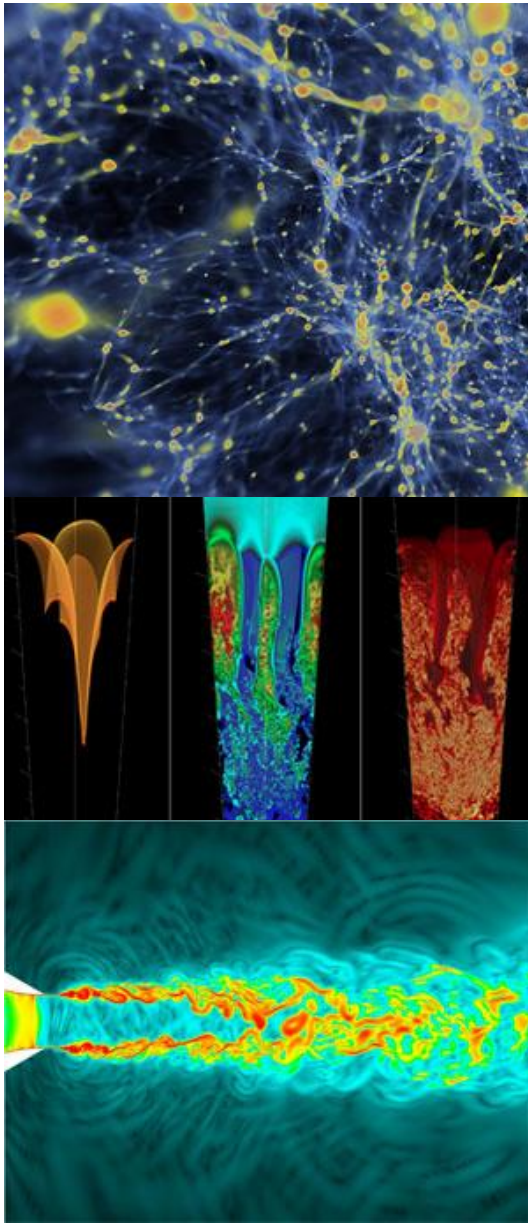
BG Partitions ("blocks")
Single Rack



- 1 I/O node for each 64 compute nodes, hardwired to specific set of 64
  - *Minimum partition size of 64 nodes*
- Partition sizes: 64, 128, 256, 512, 1024, …
  - *Any partition < 512 nodes will get a mesh network layout and not a torus.*
  - *Any partition <512 nodes will get a non-optimal I/O tree network.*
  - *Do not do performance testing on <512 nodes*
- Smaller partitions are enclosed inside of larger ones
  - *Not all partitions are available at all times*
  - *Once a job is running on one of the smaller partitions, no jobs can run on the enclosing larger partitions*
- Configuration changes frequently
  - ***partlist** shows partition state*
- Processes are spread out in a pre-defined mapping, alternate and sophisticated mappings are possible

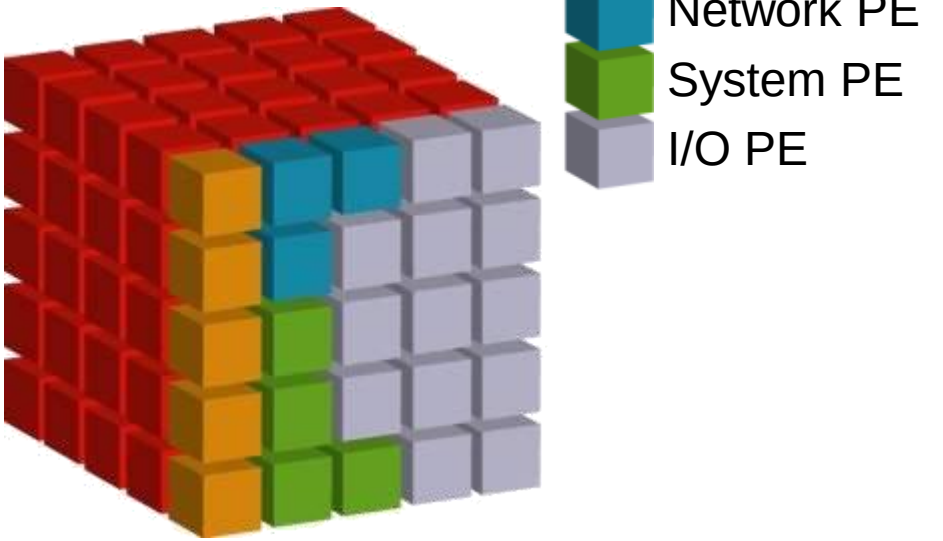
Data Analysis and Visualization

- Visualization Hardware: Eureka DA Cluster
  - 100 servers: dual quad-core Xeon processors, two Quadro FX5600 GPUs, 32 GB memory
  - Aggregate 111 TF GPU (SP), 3.2 TB memory, 312 GB GPU memory
- Production tools for high-performance visualization
  - ParaView, VisIt
- Support for users' visualization and analysis needs:
  - tools and methods for high-performance post processing of large datasets
  - interactive data exploration and batch visualization
  - production visualization



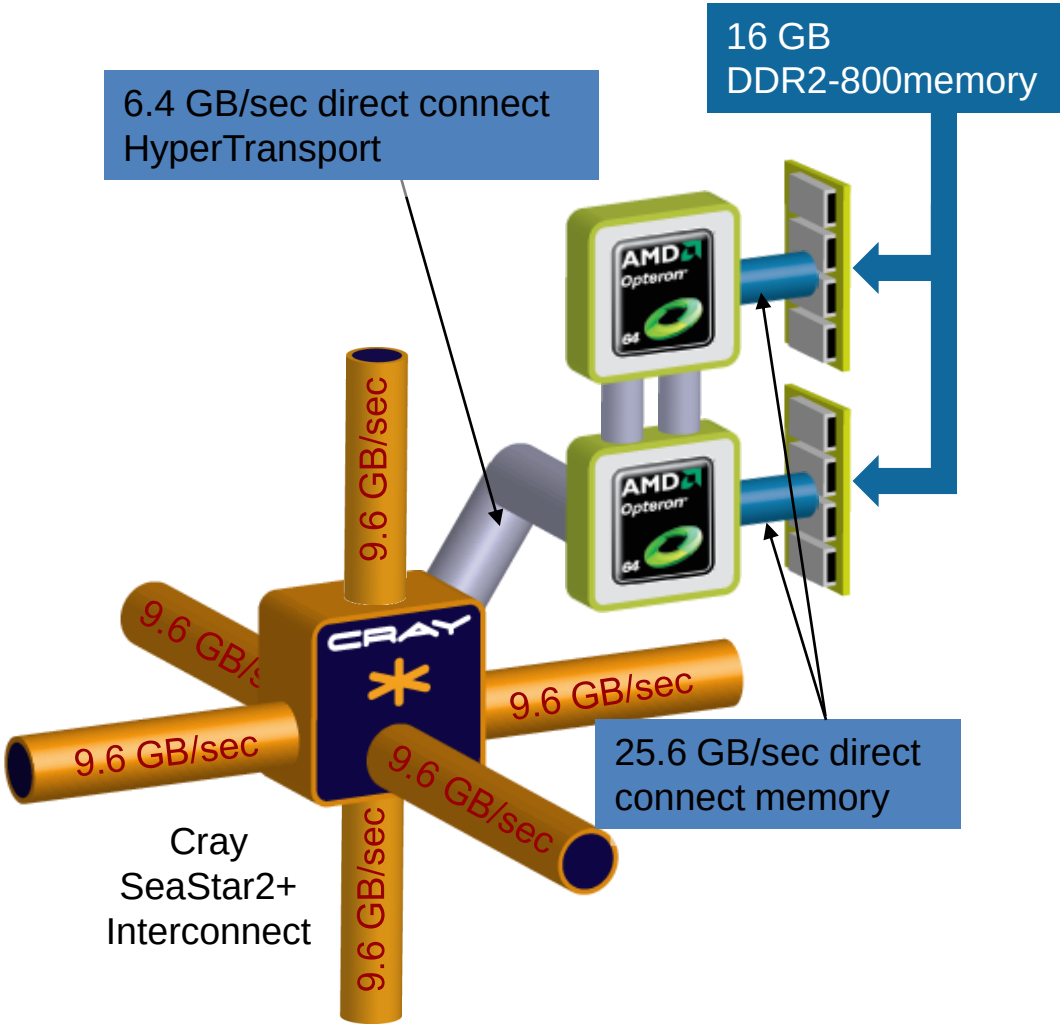
OLCF CrayXT "Jaguar" System

- Jaguar consists of an XT4 and an XT5 partition
- Ability to submit to either partition from a single set of login nodes is available



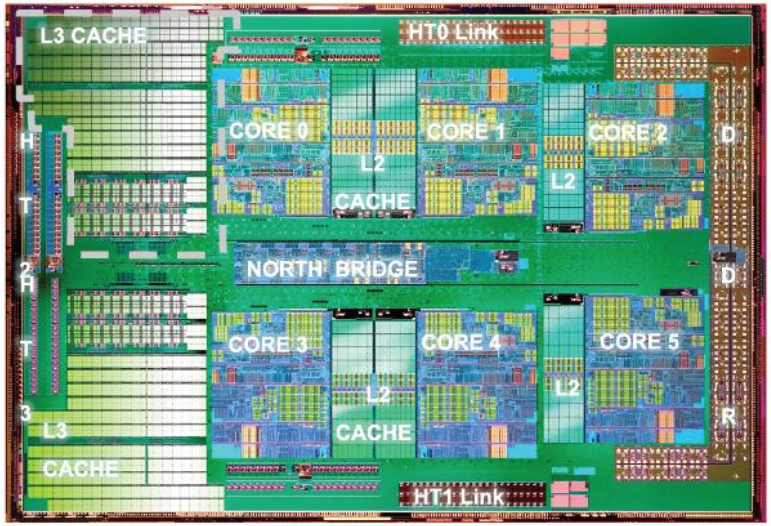
Peak performance	2.332 PF
System memory	300 TB
Disk space	10 PB
Disk bandwidth	240+ GB/s
Compute Nodes	18,688
AMD "Istanbul" Sockets	37,376
Size	5,000 feet^2
Cabinets	200 (8 rows of 25 cabinets)

XT5 Node

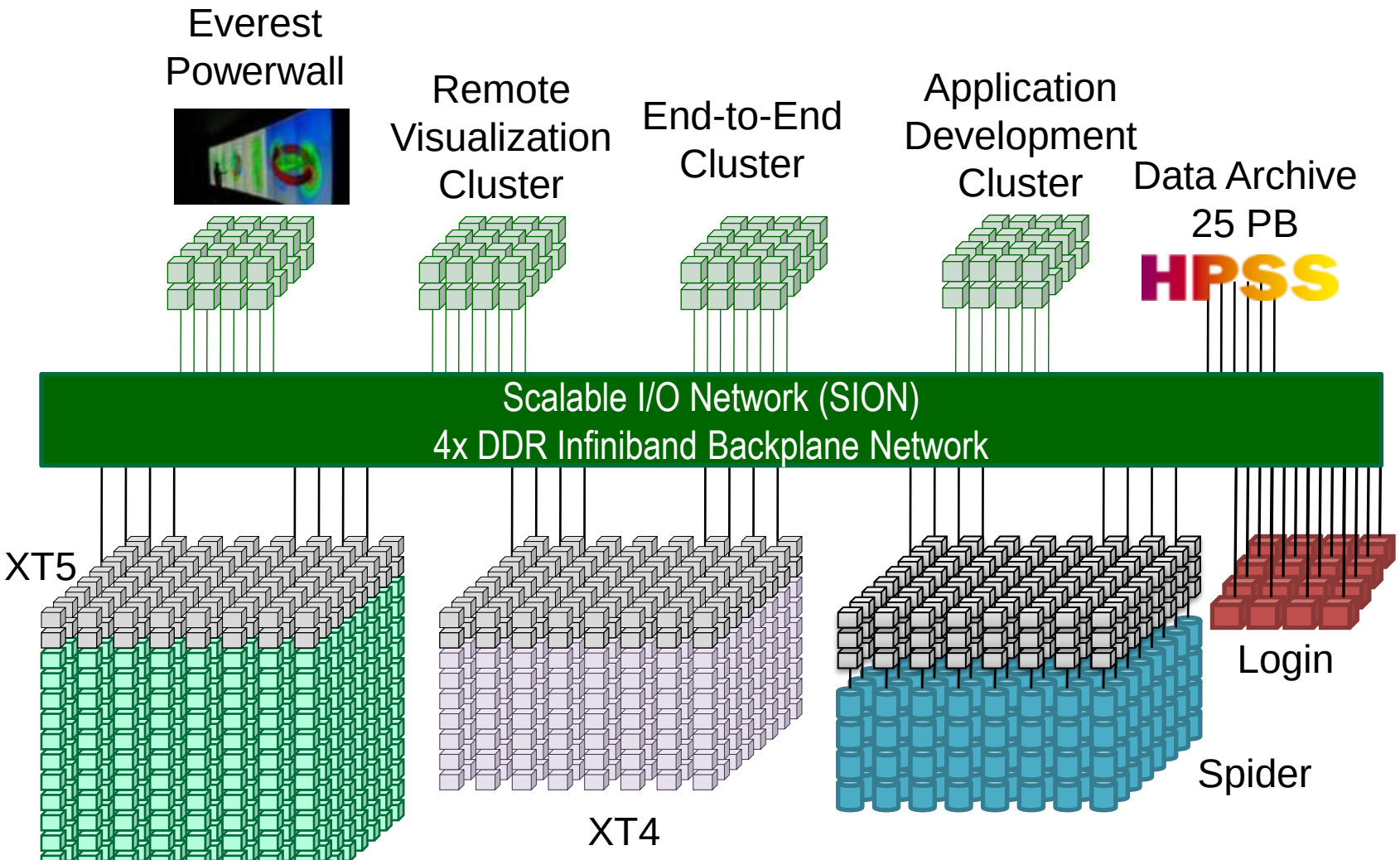


Cray XT5 Node Characteristics	
Number of Cores	12
Peak Performance	125 Gflops/sec
Memory Size	16 GB per node
Memory Bandwidth	25.6 GB/sec

AMD Opteron 2435 (Istanbul) processors



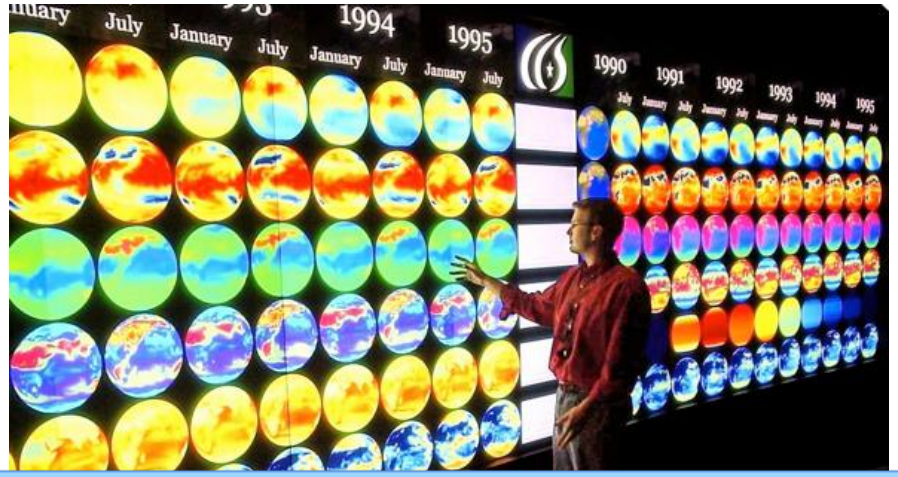
OLCF Center-wide file system: Spider



ORNL's External login nodes and shared storage provide a single entry for users into a cluster of supercomputers

OLCF Visualization Resources

- Hardware
  - Lens cluster
    - 32-node, 512-core, 2TB aggregate memory
    - 1 NVIDIA GTX8800 (768MB) & 1 Tesla C1060 (4GB) per node
  - EVEREST powerwall
    - 30’ x 8’ display with 11,520 x 3,072 resolution
- Software
    - Including VisIt, EnSight Gold (DR), ParaView, POV-Ray, AVS/Express, R MPI, IDL, VirtualGL, NX



Every INCITE project is assigned a single “visualization liaison” as a single point of contact for all post-analysis data processing

Support visualization tools	Convert data
Provide parallel data analysis support	Produce publication-ready images
Perform statistical analyses	Produce movies and animations
Research new data exploration techniques	Highlight science successes to visitors
Develop custom visualization tools and algorithms	Large display support

2012 Call for proposals

- Opens April 13, 2011 closes **June 30, 2011**
- Awards made independently of funding source
- 1.6B processor hours to be allocated for CY 2012
- Average project award to exceed **20M processor hours**
- Applicants must present evidence that their proposed production simulations can make effective use of a significant fraction, in most cases **20% or more**, of the LCF systems offered for allocation

Projects suitable for INCITE

Does your project satisfy most of the following criteria?	
✓	High-impact science and engineering with specific objectives
	AND
✓	Computationally intensive runs that cannot be done anywhere else
✓	Jobs can use at least 20% of the system
✓	Campaign requires tens of millions of CPU hours
✓	Computations are efficient on INCITE's LCF systems

Key questions to ask

- Is another resource more appropriate?
- Is both the scale of the runs **and** the time demands of the problem of LCF scale?
- Do you need specific LCF hardware?
- Do you have the people ready to do this work?
- Do you have a post-processing strategy?
- Do you have a workflow?

Some of the above characteristics are negotiable, so make sure to discuss atypical requirements with the centers

Some limitations on what can be done

- Laws regulate what can be done on these systems
  - LCF systems have cyber security plans that bound the types of data that can be used and stored on them
- Some kinds of information we cannot have
  - Personally Identifiable Information (PII)
  - Classified Information or National Security Information
  - Unclassified Controlled Nuclear Information (UCNI)
  - Naval Nuclear Propulsion Information (NNPI)
  - Information about development of nuclear, biological or chemical weapons, or weapons of mass destruction

Inquire if you are unsure or have questions

Other things we need to know

- Are you using proprietary input or software? Are you producing proprietary output?
  - Proprietary materials are things that you reserve rights to

- Does your project have export control classification?
  - Export control can be related to the topic under study (for example, fuel rods)
  - Export control can be relevant to the software application you're using

This information helps the centers determine which user agreement is appropriate and the levels of data protection needed

Proposal form: Outline

1	Principal investigator and co-principal investigators
2	Project title
3	Research category
4	Project summary (50 words)
5	Computational resources requested
6	Funding sources
7	Other high-performance computing support for this project
8	Project narrative, other materials
	(A) executive summary (*1 page*)
	(B) project narrative, including computational readiness, job characterization (*15 pages*)
	(C) personnel justification (*1 page*)
9	Application packages
10	Proprietary and sensitive information
11	Export control
12	Monitor information

Starting the proposal

- **Principal investigator, other contacts**
  - Author/PI needs to ensure that all contact information is current
  - Ask around your organization for the institutional contact (i.e., the person at your institution who can sign a user agreement)
- **Project title**
  - Pick a title you will be proud of seeing in many, many places
  - Be succinct!
- **Project summary**
  - 2 sentences suitable for the public (ex., *Science News*)

Computer resources:
Identify which system you need

- **If your project needs a single primary resource**
  - Identify the primary resource
- **If your project needs multiple primary resources**
  - Identify each of the primary resources by adding resources for the same year
  - Justify your need for and ability to effectively use multiple resources in the project narrative

	Argonne LCF	Oak Ridge LCF
System	IBM Blue Gene/P	Cray XT5
Name	Intrepid	Jaguar
Compute nodes	40,960	18,688
Processing cores	163,840	224,256
Memory, terabytes	80	>300
Peak performance, teraflops	557	2033

Computational resource request

- How many years will your project need (1–3)?
- Things that will slow you down the first year
  - Porting and code development
  - User agreements for all institutions and people involved
  - Paperwork for proprietary use
- Mind the units!
  - Processor (core) hours for system
  - Disk storage in gigabytes for both Home and Scratch space
  - Mass (tape) storage in gigabytes or terabytes (specify)

- Units are core-hours; you are charged for all cores on a node
- Intrepid: You are charged for all cores in your partition. Large partitions are in increments of a rack (1,024 nodes, 4,096 cores)
- Jaguar: Hours are core-hours, but you must request cores in increments of nodes (i.e., 12 cores)

Project narrative: Impact of the work

- **Audience**
  - Computational-science-savvy senior scientists/engineers, and faculty
  - Not everyone will be well versed in your approach
- **Story elements**
  - What the problem is, and its significance
  - Key objectives, key simulations/computations, project milestones
  - Approach to solving the problem, its challenging aspects, preliminary results
  - Impact of a successful computational campaign — the big picture
  - Reasons why it is important to carry out this work now

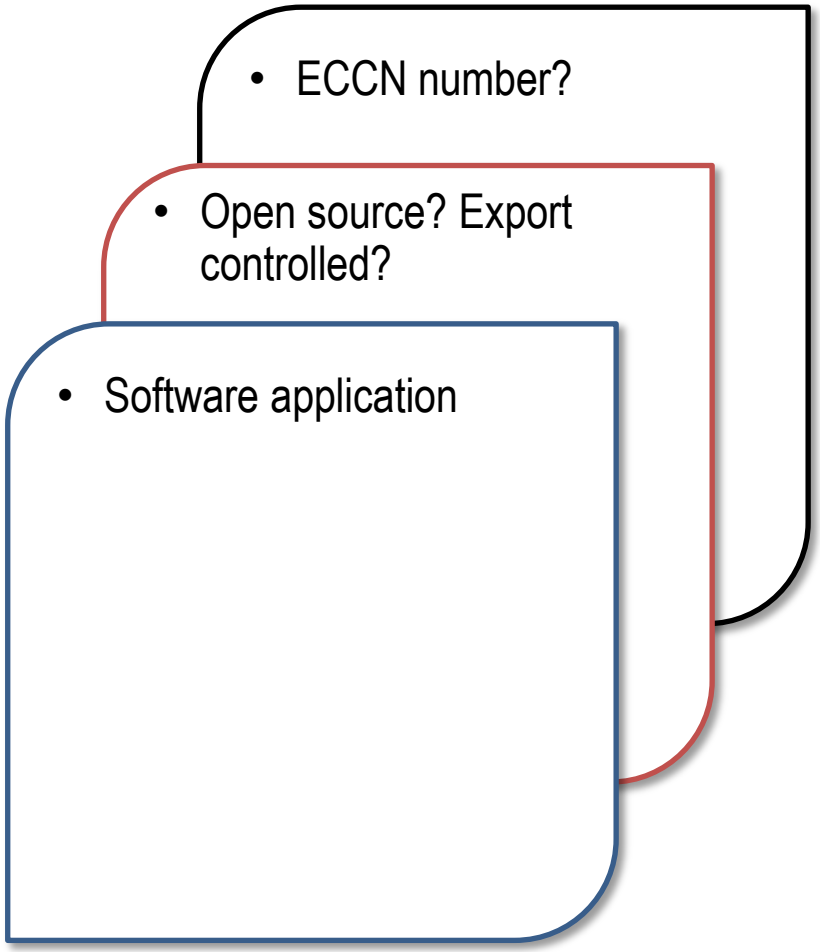
Project narrative: Computational approach

- **Experience and credibility**
  - Successful proposal teams demonstrate a clear understanding of petascale computing and can optimally use these resources to accomplish the stated scientific/technical goals
- **Programming languages, libraries and tools used**
  - Check that what you need is available on the system
- **Description of underlying formulation**
  - Don't assume reviewers know all the codes
  - Make it clear that the code you plan to employ is the correct tool for your research plan
  - If you plan to use a private version of a well-known code, delineate the differences

Application packages

List all software application packages/suites to be used (Note: Long lists may reduce credibility)

- What will be used to set up computations?
- What are the codes for the main simulation/modeling?
- What will be used to analyze results?



Application credentials

- **Port your code before submitting the proposal**
  - Check to see if someone else has already ported it
  - Request a startup account if needed
- **It is very hard to embarrass a >150,000-processor system**
  - Prove application scalability in your proposal
  - Run example cases at full scale
  - Make examples as close to production runs as possible
  - If you cannot show proof of runs at full scale, then provide a very tight story about how you will succeed

Discretionary time on Jaguar and Intrepid can be requested now, for benchmarking

Computational Campaign

- **Describe the kind of runs you plan with your allocation**
  - L exploratory runs using M nodes for N hours
  - X big runs using Y nodes for Z hours
  - P analysis runs using Q nodes for R hours
- **Big runs often have big output**
  - Show you can deal with it and understand the bottlenecks
  - Understand the size of results, where you will analyze them, and how you will get the data there

A sample of codes with local expertise available at Argonne and Oak Ridge

Application	Field	ALCF	OLCF
FLASH	Astrophysics	✓	✓
MILC,CPS	LQCD	✓	✓
Nek5000	Nuclear energy	✓	
Rosetta	Protein structure	✓	
DCA++	Materials science		✓
ANGFMC	Nuclear structure	✓	
NUCCOR	Nuclear structure		✓
Qbox	Chemistry	✓	✓
LAMMPS	Molecular dynamics	✓	✓
NWChem	Chemistry	✓	✓
GAMESS	Chemistry	✓	✓
MADNESS	Chemistry	✓	✓
CHARMM	Molecular dynamics	✓	✓
NAMD	Molecular dynamics	✓	✓

Application	Field	ALCF	OLCF
AVBP	Combustion	✓	
GTC	Fusion	✓	✓
Allstar	Life science	✓	
CPMD, CP2K	Molecular dynamics	✓	✓
CCSM3	Climate	✓	✓
HOMME	Climate	✓	✓
WRF	Climate	✓	✓
Amber	Molecular dynamics	✓	✓
enzo	Astrophysics	✓	✓
Falkon	Computer science/HTC	✓	✓
s3d	Combustion		✓
DENOVO	Nuclear energy		✓
LSMS	Materials science		✓
GPAW	Materials science	✓	

Programming models

- **Parallel Programming System**
  - MPI (MPICH2) is the work horse on both platforms
    - Cray MPI based on MPICH2
    - ARMCI/Global Arrays is available
  - OpenMP on nodes
  - Some groups have “rolled their own” at lower level, e.g., QCD
- **Special needs? Inquire**
  - e.g., Python, custom kernel

OLCF-supported Tools/Libraries/Packages

- Analysis
  - Matlab
  - NCO
  - UDUNITS
  - Ferret
  - gnuplot
  - grace
  - IDL
  - ncl
  - ncview
  - PGPLOT
  - VisIt
  - ParaView
  - Ensight
- Libraries
  - Communication
    - BLACS
    - Global Arrays
  - I/O
    - HDF5
    - netcdf
    - pnetcdf
    - Szip
    - silo
  - Math
    - acml
    - ARPACK
    - ATLAS
    - Aztec
    - BLAS
    - fftpack
    - fftw
    - gsl
    - hypre
    - LAPACK
    - Cray libsci
    - METIS
    - MUMPS
    - ParMETIS
    - PETSc
    - pspline
    - ScaLAPACK
    - sprng
    - Sundials
    - SuperLU
    - Trilinos
    - UMFPACK
- Program Dev
  - cmake
  - doxygen
  - git
  - mercurial
  - subversion
  - Debugging
    - Totalview
    - Valgrind
    - DDT
  - Performance
    - Apprentice2
    - craypat
    - fpmpi
    - gptl
    - mpe2
    - mpip
    - PAPI
    - TAU
    - vampir

ALCF-supported Tools/Libraries/Packages

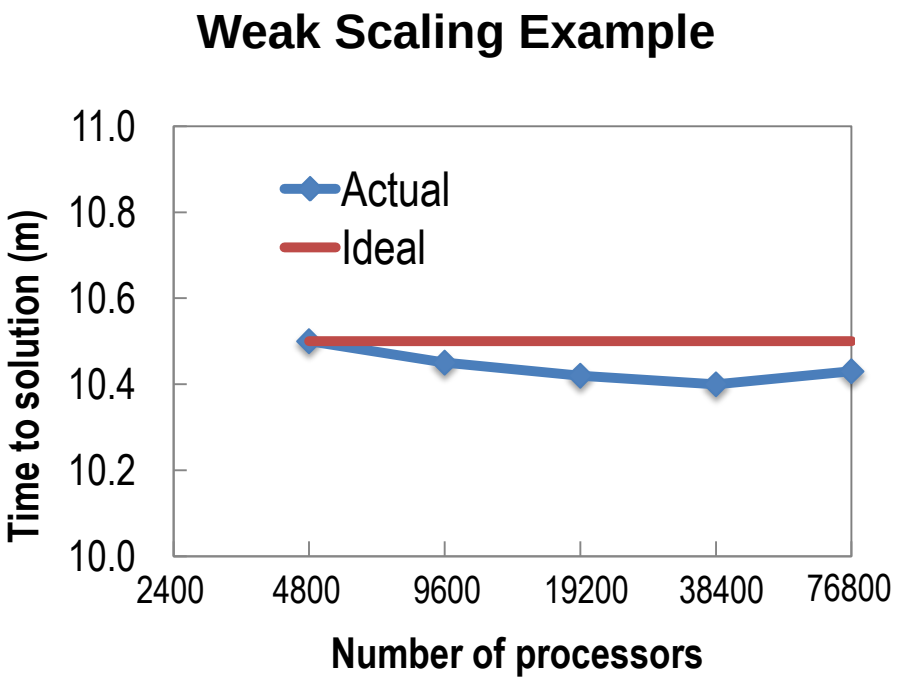
- Analysis
  - Gnuplot
  - Paraview
  - Visit
- Libraries
  - Communication
    - BLACS
    - DCMF
    - Global Arrays
  - I/O
    - HDF5
    - netcdf
    - pnetcdf
    - szip
  - Math
    - ATLAS
    - BLAS/ GotoBLAS
    - essl
    - Fftpack
    - Fftw
    - p3dfft
    - hypre
    - LAPACK
    - Mass/massv
    - METIS
    - MUMPS
    - ParMETIS
    - PETSc
    - SCALAPACK
    - Spooles
    - SuperLU
    - Trilinos
- Program Dev
  - cmake
  - Doxygen
  - git
  - mercurial
  - subversion
  - Debugging
    - gdb
    - Totalview
    - Valgrind
    - coreprocessor
  - Performance
    - Darshan
    - HPC Toolkit (IBM)
    - HPCToolkit
    - (Rice)
    - Fpmpi
    - mpe2
    - mpip
    - PAPI
    - TAU
    - UPC (universal performance counter)
- Other
  - ZeptoOS

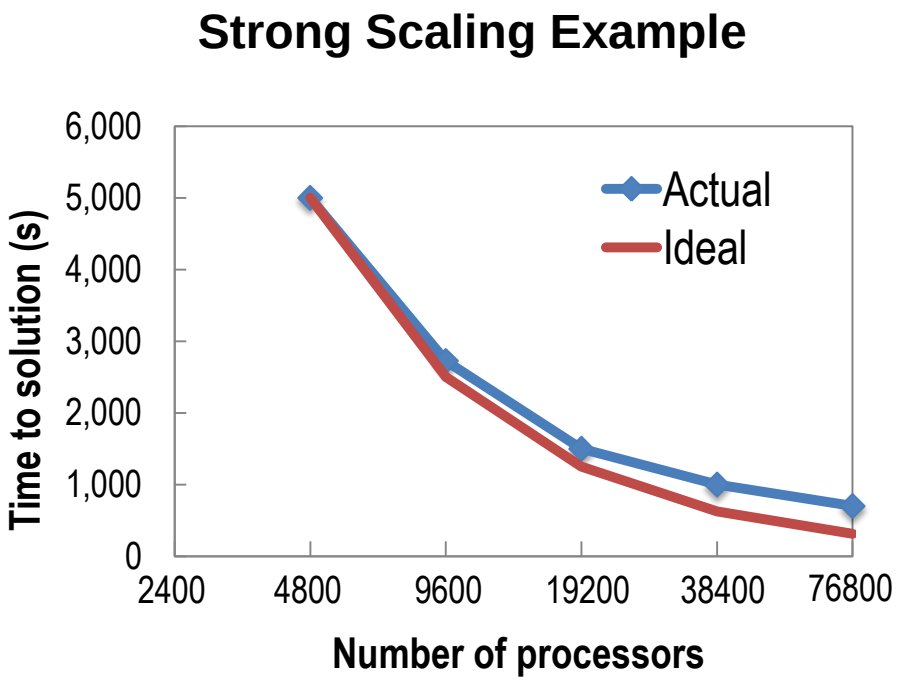
If a you have a requirement that is not listed, ask!

Parallel performance: Direct evidence

WEAK SCALING DATA	STRONG SCALING DATA
Increase problem size as resources are increased	Increase resources (nodes) while doing the same computation

Pick the approach(es) relevant to your work and show results





Parallel performance: Direct evidence

- **Performance data should support the required scale**
  - Use similar problems to what you will be running
  - Show that you can get to the range of processors required
  - Best to run on the same machine, but similar size runs on other machines can be useful
- **Describe how you will address any scaling deficiencies**
  - Be aware of scaling data from other groups and literature

Consider applying for LCF Discretionary time

- Requests may be submitted anytime throughout the year
- Awards of time on the order of 1M processor hours

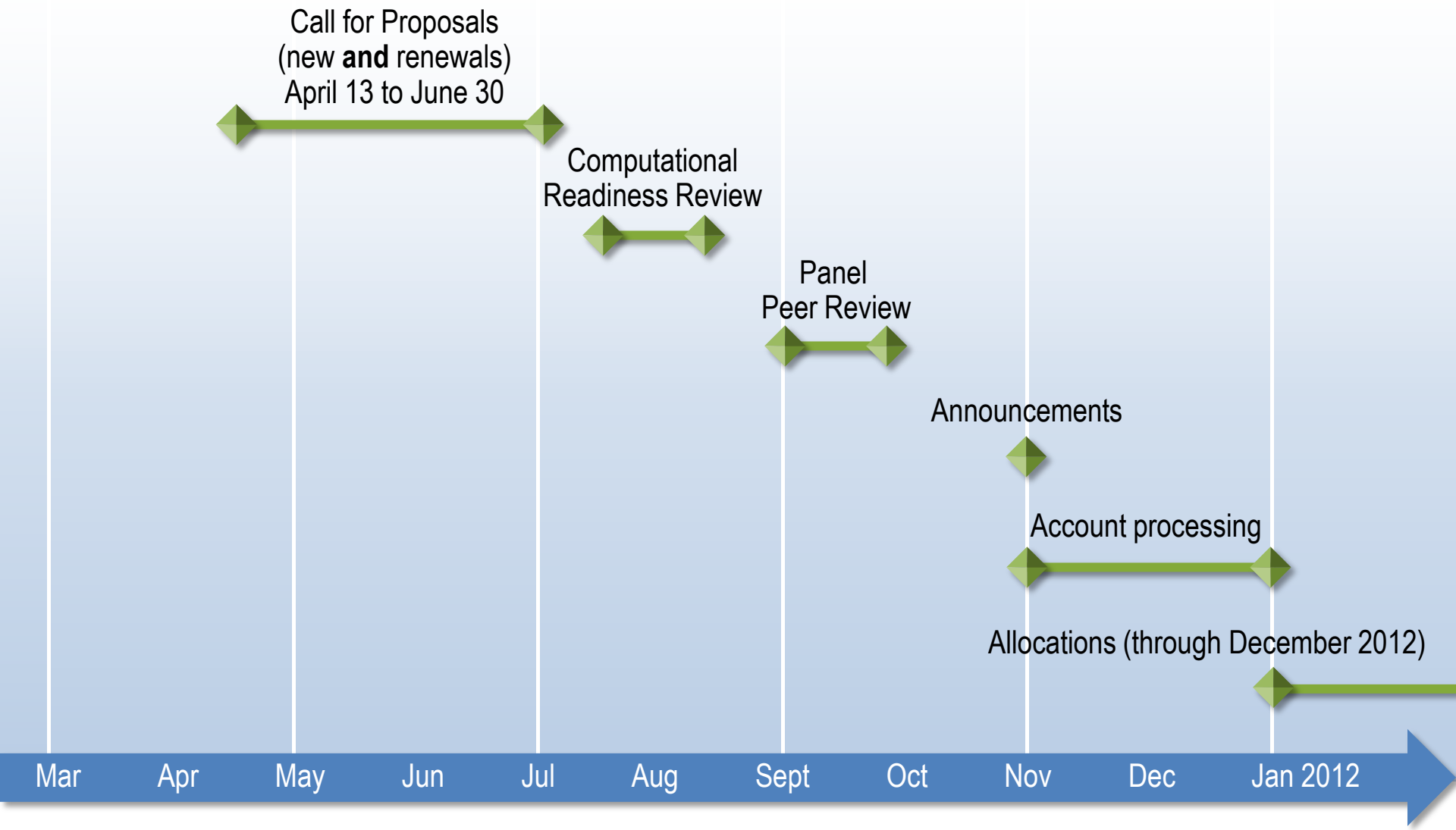
I/O Requirements

- Restart I/O - Application initiated program restart data
  - I/O technique used, e.g., MPI I/O, HDF5, raw
  - Number of processors doing I/O, number of files
  - Sizes of files and overall dump
  - Periodicity of the checkpoint process
- Analysis I/O - Application written files for later analysis
  - I/O technique used, e.g., pNetCDF, pHDF5
  - Number of processors doing I/O, number of files
  - Sizes of files and overall dump
- Archival I/O - Data archived for later use/reference
  - Number and sizes of files
  - Retention length
  - If archived remotely, the transport tool used, e.g., GridFTP

Final checks

- You may save your proposal at any time without having the entire form complete
- Your Co-PIs may also log in and edit your proposal
- Required fields must be completed for the form to be successfully submitted
  - An incomplete form may be saved for later revisions
- After submitting your proposal, you will not be able to edit it

2012 INCITE: Schedule for proposals



Twofold review process

	New proposal assessment	Renewal assessment
1 **Peer review: INCITE panels**	• Scientific and/or technical merit • Appropriateness of proposal method, milestones given • Team qualifications • Reasonableness of requested resources	• Change in scope • Met scientific milestones • On track to meet future milestones
2 **Computational readiness review: LCF centers**	• Appropriateness for requested resources and computational approach • Technical readiness	• Met technical/ computational milestones • On track to meet future milestones

Award announcements

- Notice comes from INCITE Manager, Julia White, in November 2011
- Welcome and startup information comes from centers
  - Agreements to sign: Start this process as soon as possible!
  - How to get accounts
- User Assistance is geared to help you succeed
- Centers provide expert-to-expert assistance to help you get the most from your allocation
  - Scientific “Liaisons” and “Catalysts” (OLCF / ALCF)

PI obligations

Let us know your achievements and challenges

- Provide quarterly status updates (on supplied template)
  - Milestone reports
  - Publications, awards, journal covers, presentations, etc., related to the work
- Provide highlights on significant science/engineering accomplishments as they occur
- Submit annual renewal request
- Complete annual surveys
- Encourage your team to be good citizens on the computers
- Use the resources for the proposed work

It is a small world…

- INCITE program and center resources will continue to grow as researchers around the world require larger systems for high-impact results
- Let the science agency that funds your work know how significant the INCITE program and the Leadership Computing Facilities will be to your work
- Contact us if you have questions: we want to hear from you

Relevant links

INCITE Program
http://www.doeleadershipcomputing.org/

Argonne Discretionary Program
https://wiki.alcf.anl.gov/index.php/Discretionary_Allocations

Oak Ridge Discretionary Program
http://www.nccs.gov/user-support/access/project-request/

Contact the center if you'd like to request Discretionary time on Intrepid or Jaguar for benchmarking

Contacts

For details about the INCITE program:

www.doeleadershipcomputing.org
INCITE@DOEleadershipcomputing.org

For details about the centers:

www.olcf.ornl.gov
help@nccs.gov, 865-241-6536

www.alcf.anl.gov
support@alcf.anl.gov, 866-508-9181